

链家武汉二手房数据分析

李倩

2023-10-20

数据介绍

本报告链家数据获取方式如下：

报告人在 2023 年 9 月 12 日获取了链家武汉二手房网站数据。

- 链家二手房网站默认显示 100 页，每页 30 套房产，因此本数据包括 3000 套房产信息；
- 数据包括了页面可见部分的文本信息，具体字段及说明见作业说明。

说明：数据仅用于教学；由于不清楚链家数据的展示规则，因此数据可能并不是武汉二手房市场的随机抽样，结论很可能有很大的偏差，甚至可能是错误的。

```
## Rows: 3,000
## Columns: 18
## $ property_name      <chr> "南湖名都A区", "万科紫悦湾", "东立国际", "新都汇", "~
## $ property_region    <chr> "南湖沃尔玛", "光谷东", "二七", "光谷广场", "团结大~
## $ price_ttl           <dbl> 237.0, 127.0, 75.0, 188.0, 182.0, 122.0, 99.0, 193.8~
## $ price_sqm           <dbl> 18709, 14613, 15968, 15702, 17509, 10376, 12346, 163~
## $ bedrooms           <dbl> 3, 3, 1, 3, 3, 3, 2, 3, 4, 3, 5, 3, 4, 3, 3, 2, 3, 4~
## $ livingrooms        <dbl> 1, 2, 1, 2, 2, 2, 1, 2, 1, 2, 2, 2, 2, 1, 2, 2, 2, 2~
## $ building_area       <dbl> 126.68, 86.91, 46.97, 119.73, 103.95, 117.59, 80.19, ~
## $ directions1        <chr> "南", "南", "南", "北", "东南", "南", "南", "南", "~
```

```
## $ directions2      <chr> "北", NA, NA, "东", NA, "北", NA, "北", "北", "北", ~
## $ decoration        <chr> "精装", "精装", "简装", "精装", "简装", "精装", "简~
## $ property_t_height <dbl> 17, 28, 18, 32, 34, 34, 7, 34, 5, 7, 25, 32, 8, 31, ~
## $ property_height   <chr> "中", "中", "低", "高", "中", "低", "低", "中", "低"~
## $ property_style    <chr> "塔楼", "板楼", "塔楼", "塔楼", "板塔结合", "板楼", ~
## $ followers         <dbl> 3, 1, 3, 2, 3, 1, 0, 0, 2, 0, 0, 0, 10, 0, 0, 1, 0, ~
## $ near_subway       <chr> "近地铁", NA, "近地铁", "近地铁", NA, NA, "近地铁", ~
## $ if_2y             <chr> NA, "房本满两年", NA, "房本满两年", "房本满两年", "~
## $ has_key           <chr> "随时看房", "随时看房", "随时看房", "随时看房", "随~
## $ vr               <chr> NA, "VR看装修", NA, NA, "VR看装修", NA, "VR看装修", ~

##   property_name      property_region      price_ttl      price_sqm
## Length:3000      Length:3000      Min.    : 10.6      Min.    : 1771
## Class :character  Class :character  1st Qu.: 95.0      1st Qu.:10799
## Mode  :character  Mode  :character  Median : 137.0     Median :14404
##                                     Mean  : 155.9      Mean   :15148
##                                     3rd Qu.: 188.0     3rd Qu.:18211
##                                     Max.   :1380.0     Max.   :44656

##   bedrooms      livingrooms      building_area      directions1
## Min.    :1.000      Min.    :0.000      Min.    : 22.77      Length:3000
## 1st Qu.:2.000      1st Qu.:1.000      1st Qu.: 84.92      Class :character
## Median :3.000      Median :2.000      Median : 95.55      Mode  :character
## Mean    :2.695      Mean    :1.709      Mean    :100.87
## 3rd Qu.:3.000      3rd Qu.:2.000      3rd Qu.:117.68
## Max.    :7.000      Max.    :4.000      Max.    :588.66

##   directions2      decoration      property_t_height      property_height
## Length:3000      Length:3000      Min.    : 2.00      Length:3000
## Class :character  Class :character  1st Qu.:11.00      Class :character
## Mode  :character  Mode  :character  Median :27.00      Mode  :character
##                                     Mean    :24.22
##                                     3rd Qu.:33.00
##                                     Max.    :62.00

##   property_style      followers      near_subway      if_2y
## Length:3000      Min.    : 0.000      Length:3000      Length:3000
```

```
## Class :character 1st Qu.: 1.000 Class :character Class :character
## Mode :character Median : 3.000 Mode :character Mode :character
##
## Mean : 6.614
##
## 3rd Qu.: 6.000
##
## Max. :262.000
## has_key vr
## Length:3000 Length:3000
## Class :character Class :character
## Mode :character Mode :character
##
##
##

## # A tibble: 4 x 2
## decoration 数量
## <chr> <int>
## 1 其他 173
## 2 毛坯 436
## 3 简装 634
## 4 精装 1757

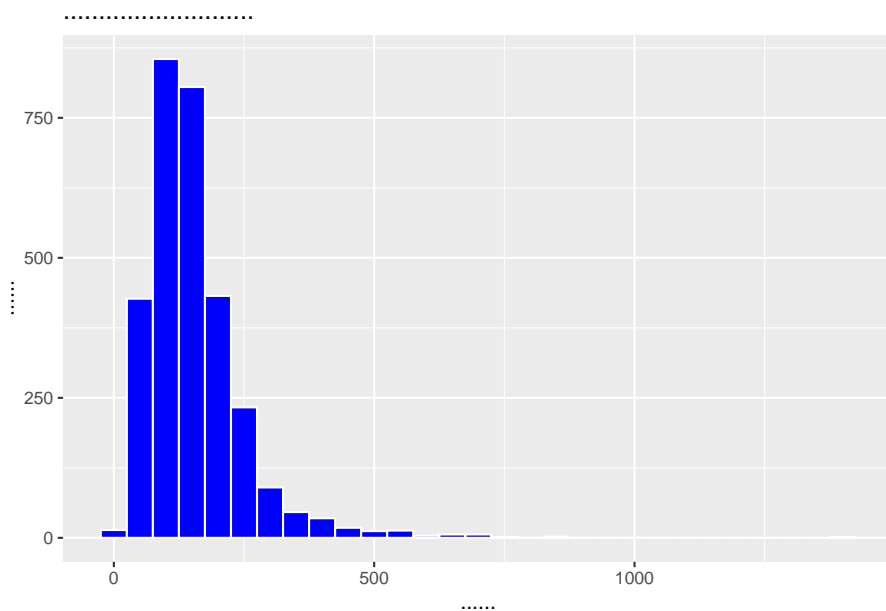
## # A tibble: 5 x 2
## property_style 数量
## <chr> <int>
## 1 塔楼 527
## 2 平房 5
## 3 暂无数据 72
## 4 板塔结合 615
## 5 板楼 1781
```

可以看到: 1.【followers】房屋的最高关注人数有 262, 最低关注人数为 0, 平均关注人数 6.614, 中位数为 3; 2.【price_ttl】房屋总价最高为 1380 万元, 最低为 10.6 万元, 平均总价为 155.9 万元, 中位数为 137 万元; 3.【price_sqm】房屋单价最高为 44656 元, 最低为 177 元, 平均单价 15148 元, 中位数为 14404 元; 4.【bedrooms】房间数最多为 7 个, 最少 1 个, 平均拥有 2.695

个房间；5. **【livingrooms】** 客厅数量最多有 4 个，最少 0 个，平均拥有 1.709 个；6. **【building_area】** 建筑面积最大为 588.66，最少为 22.77，平均面积 100.87；7. **【property_t_height】** 楼栋总层数最高为 62 层，最低 2 层，平均 24.22 层；8. **【decoration】** 房屋装修情况统计中，精装的房屋数量最多，共 1757；9. **【property_style】** 建筑形式统计中，板楼数量最多，平房数量最少。

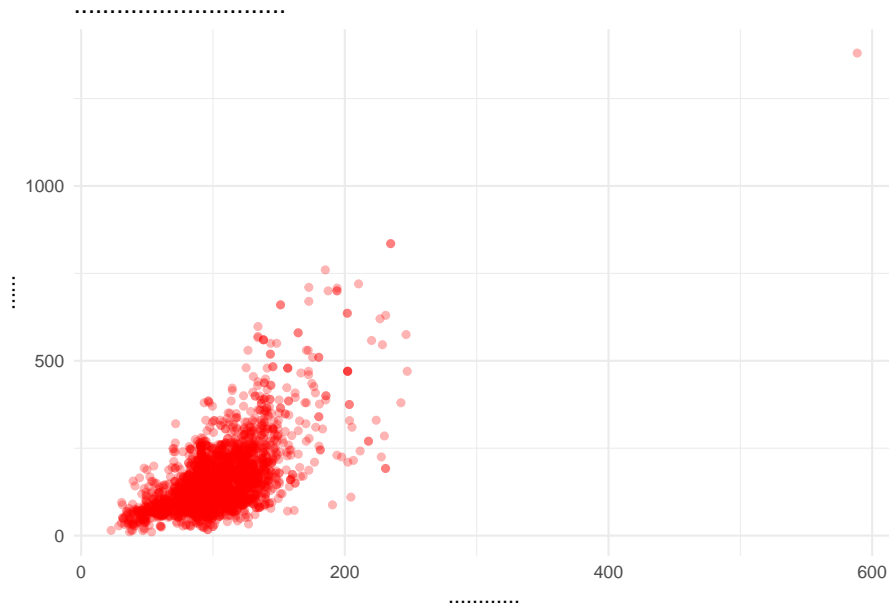
探索性分析

数值描述与图形 1 ## 房价分布



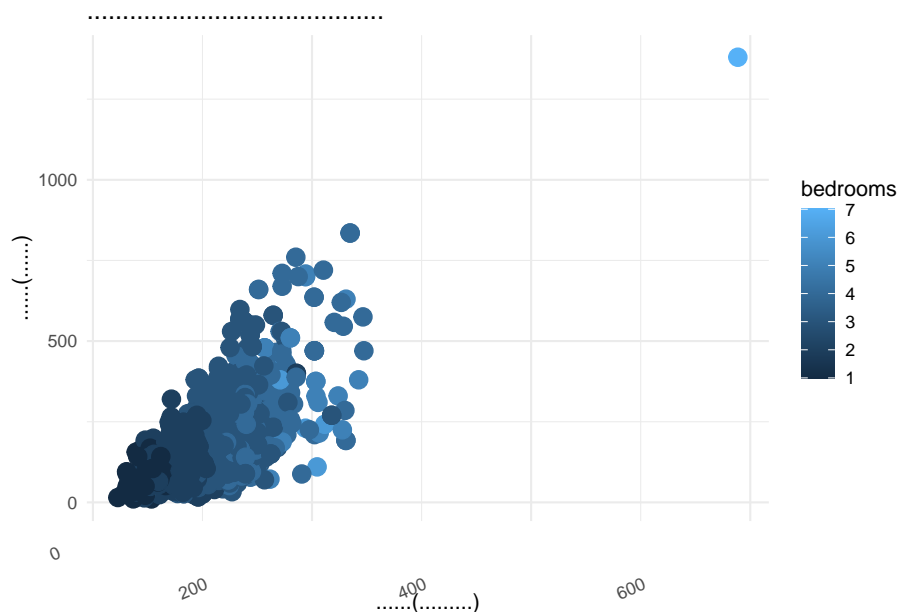
发现：1. 大部分的房屋价格集中在 0 至 400 的范围内。2. 房屋价格在大约 100 至 250 的范围内的频率是最高的，意味着这个价格区间的房屋数量最多。3. 随着价格的上升，房屋数量逐渐减少，表现为长尾分布。特别是在 400 之后，房屋数量急剧减少。4. 在 600 至 800 的价格范围内，仍然存在少量的高价房屋。

数值描述与图形 2 ## 绘制房价与房屋面积之间的散点图



发现：1. 散点图的总体趋势表明，随着房屋面积的增加，房价通常也会增加；2. 在 40-120 的房屋面积范围内，大多数的点集中在较低的房价范围，特别是在 200 以下；3. 当房屋面积超过 120 之后，房价的分布相对比较分散，但总体上还是呈上升趋势；4. 存在一些异常值或离群点，例如在房屋面积较小的时候，房价却较高，或者在房屋面积较大的时候，房价却相对较低。这可能是由于其他因素导致的，如房屋的位置、装修程度、所在的小区等。总之，虽然房屋面积与房价之间存在正相关关系，但还有其他因素可能影响房价。

数值描述与图形 3 ## 绘制房价与房屋面积/房间数之间关系的散点图



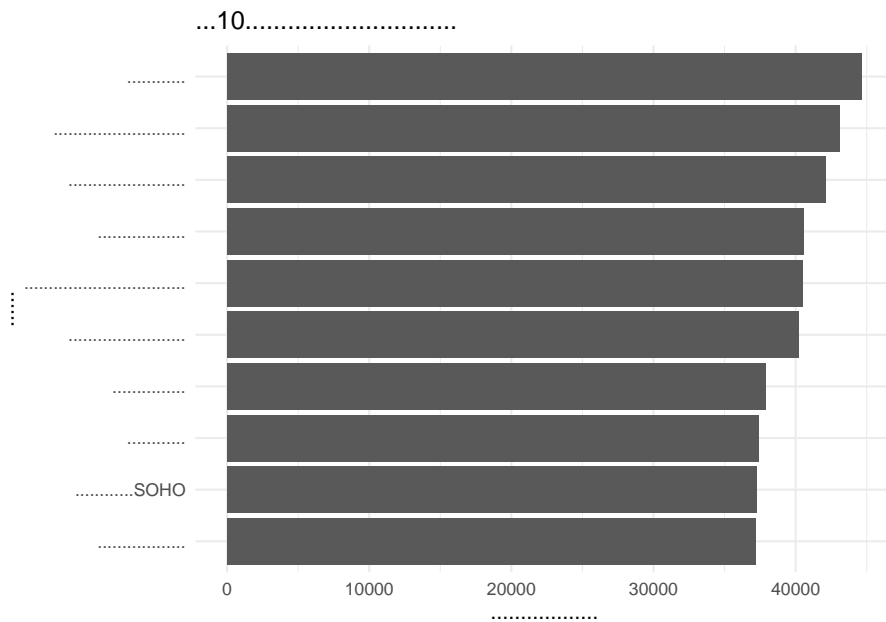
发现：1. 总价与面积关系：房屋的总价与其面积呈现出明显的正相关关系。即随着面积的增加，房屋的总价也在上升。2. 分布特点：大部分的数据点集中在左下角，表示大多数的房屋面积和总价都相对较低。但是，也有一些数据点散布在右上角，表示有些房屋即使面积不大，但总价也很高，可能这些房屋位于高价地段或有其他特殊属性。3. 离群点：在图的右侧和上方，我们可以看到一些离群点。这些可能代表特殊的房屋，例如，面积不大但总价很高的房屋，或者面积很大但总价相对较低的房屋。需要进一步分析这些房屋的其他属性，才能判断其总价的形成原因。

探索问题 1 房屋单价较高的小区是否均在繁华地带？# 房屋单价排名前 10 的小区

A tibble: 10 x 2

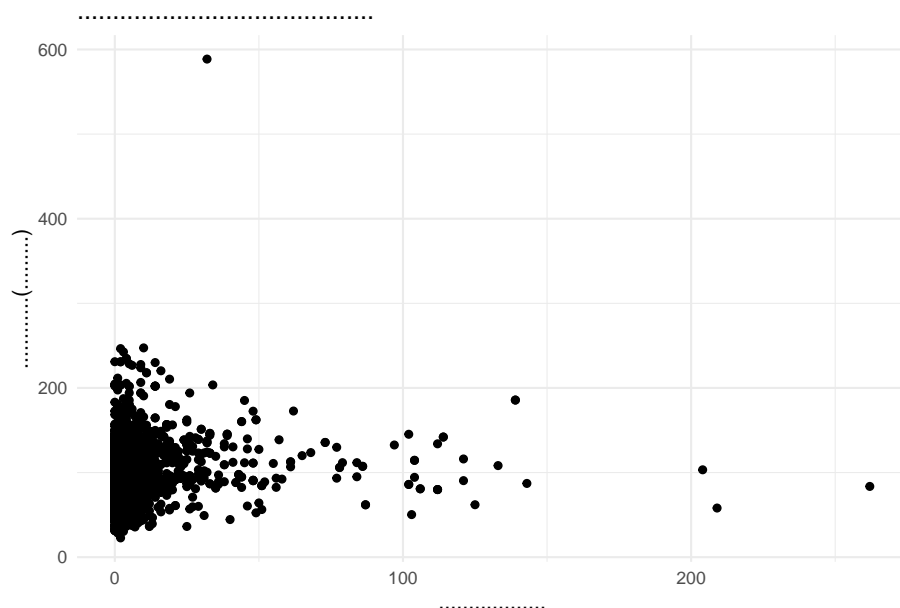
	property_name	average_price
	<chr>	<dbl>
## 1	中商宿舍	44656
## 2	复地东湖国际五六期	43094
## 3	复地东湖国际一期	42130.
## 4	华发外滩首府	40594

##	5	中信泰富滨江金融城九悦	40492
##	6	复地东湖国际二期	40180.
##	7	华发中城荟	37910.
##	8	金都汉宫	37366
##	9	中央花园SOHO	37247
##	10	湖北建材宿舍	37160



发现：1. 房价排名前 10 的小区基本均在热门商圈，处于繁华地带；2. 同一小区不同房子的单价也有区别。

探索问题 2 房屋面积对房屋关注度的影响



发现: 1. 主要的数据集中: 大部分关注人数集中在 0 到 25 范围内, 房屋面积分布在 40 到 160 之间。2. 缺乏明确趋势: 在关注人数较低的范围, 尽管有大量数据点, 但它们在房屋面积轴上分布较为均匀, 没有明显的线性关系。3. 孤立数据点: 关注人数 25 到 100 之间有一些分散的数据点, 它们在图上显得相对孤立, 可能代表一些特定的房屋或区域。4. 大面积房屋的关注度: 在房屋面积高于 140 的部分, 只有少数几个数据点, 这可能意味着大面积的房屋关注度不高。5. 中等关注度的房屋较少: 在关注人数 25 到 50 范围内, 数据点较少, 这可能表示这一区域的房屋表现比较平庸。6. 较低关注度的房屋面积分布广泛: 在关注人数低于 25 的部分, 房屋面积分布从 40 到 160 不等, 说明在这个关注度范围内, 房屋面积分布差异明显。这张图展示了房屋关注度与房屋面积之间的分布, 但两者之间没有明显的趋势关系。

发现总结影响武汉二手房房价的因素多种多样, 可以主要归纳为以下几点: 1. 地理位置: 这是影响房价的重要因素, 核心区域的二手房价格通常较高。这些区域通常拥有更完善的设施和更便利的交通, 因此更吸引购房者。2. 交通便利性: 房屋周边的交通便利性也会对房价产生重要影响。是否临近地铁, 交通便利的区域通常更吸引购房者, 因此房价相对较高。3. 房屋的品质: 房屋的品质和装修情况也是影响房价的因素。一些装修较好、品质较高

的房屋可能会吸引更多的购房者，从而推高房价。4. 区域特性：不同区域的二手房价格也可能存在差异。例如，武昌地区的房价相对较高，这主要受到该区域教育资源丰富、商业发达、环境优美等因素的影响。而汉口地区由于交通便利性较好，也吸引了众多购房者。这些因素可能单独或共同作用，影响武汉二手房的房价。需要注意的是，房价会随着时间和市场环境而波动，因此购房者在购买二手房时，应充分考虑这些因素，合理评估自己的需求和能力，做出明智的决策。